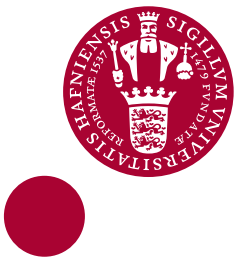


NMR spectra database for on-flight identification of HPLC-SPE-NMR data: proof of concept

Johansen, Kenneth

Publication date:
2010

Document version
Early version, also known as pre-print

FACULTY OF PHARMACEUTICAL SCIENCES
UNIVERSITY OF COPENHAGEN

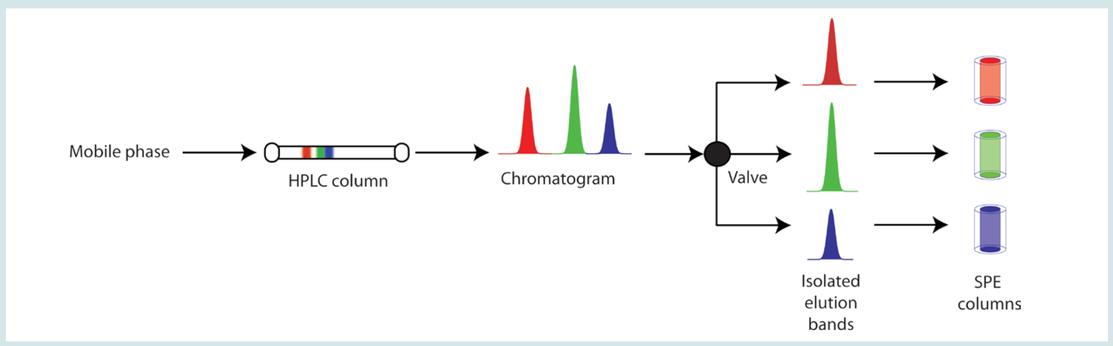
NMR spectra database for on-flight identification of HPLC-SPE-NMR data: Proof of concept

Kenneth Johansen[1*], Manfred Spraul[2], Jerzy Jaroszewski[1]
[1]Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, Universitetsparken 2, DK-2100
[2]NMR applications Group, Bruker Biospin GmbH, Silberstreifen 4, DE-76287
*E-mail: kj@farma.ku.dk

Introduction

Coupling of a separation technique to an analytical detector is commonly used in analytical chemistry. This is also referred to as a hyphenated technique and can be exemplified by e.g., GC-MS.

Advances within the area of NMR hyphenation have led to implementation of an on-line solid phase extraction (SPE) step between the outlet of the diode array detector (DAD) and the NMR where the metabolites are trapped on SPE cartridges as depicted below.



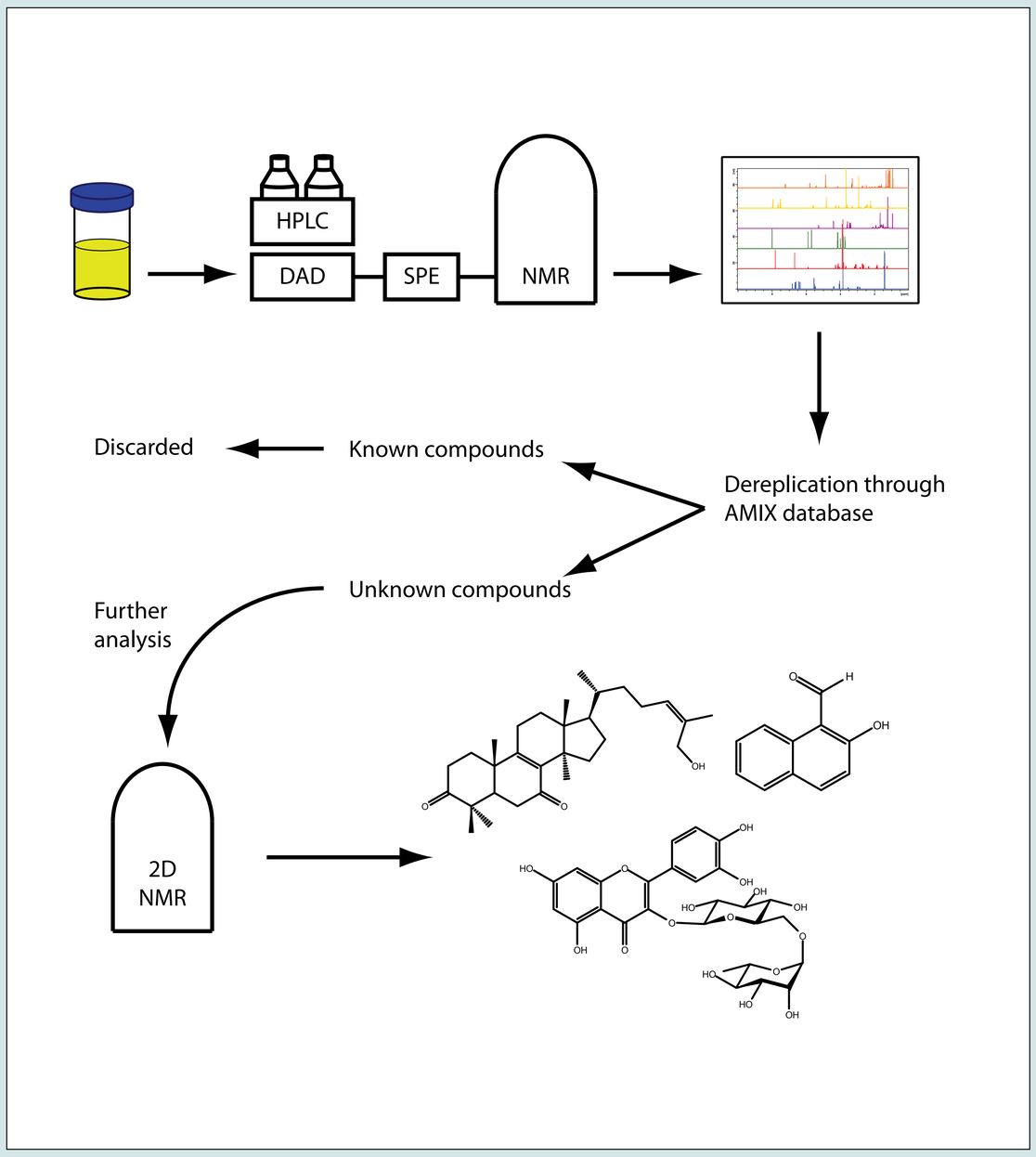
The advantages of this setup over other hyphenated NMR techniques are numerous as discussed on several occasions [1][2]. One of the major advantages is an increased signal to noise ratio (S/N) due to the possibility of doing multiple trappings. The S/N can be further enhanced by eluting into 1.7 mm NMR tubes and cryogenically cool the NMR probe as done in our lab. The major challenge with an automated system that allows analysis of an entire extract, including even minor metabolites, is the vast amount of data acquired.

Since structure elucidation of novel compounds is yet to be automated, this exercise proves to be a bottleneck which calls for a dereplication-tool to help us concentrate on the unknown compounds. Traditionally the commonly used dereplication tool is the comparably more sensitive although less information-rich mass spectrometer. With the sensitivity increase obtained within NMR we can now use the structural information from the ^{1}H spectrum for the exclusion process.

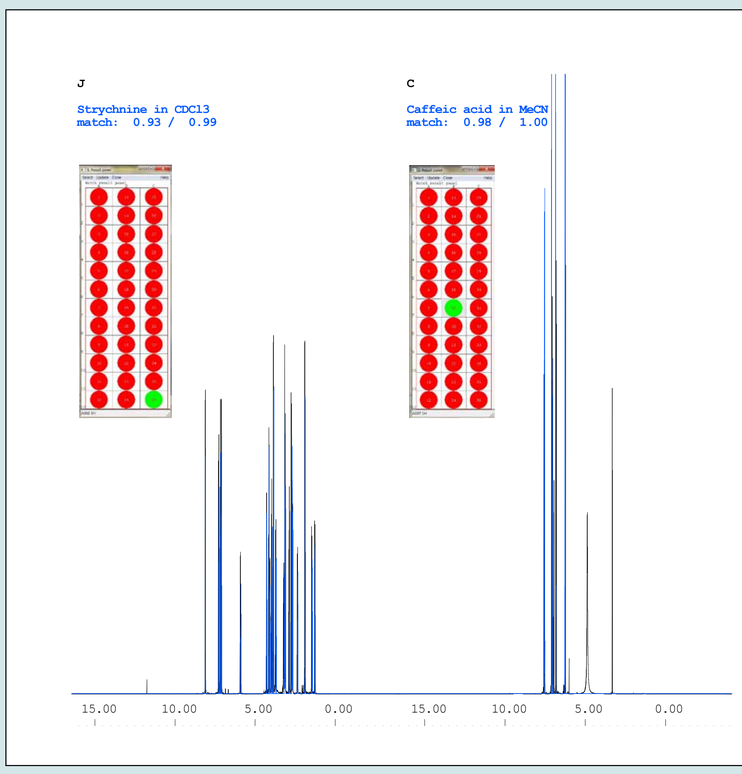
Requirements of a ^{1}H spectra database

The spectra database is build on the AMIX platform by Bruker Biospin and contains NMR data of *pure, known compounds*. To ensure that available computer capacity is only used on matching against signals instead of noise, impurities and areas containing nothing but noise are replaced by zeros. To obtain these cleaned spectra one can either clean the original spectrum through various manipulations in AMIX or by simulating the spectrum in PERCH and conduct total lineshape fitting.

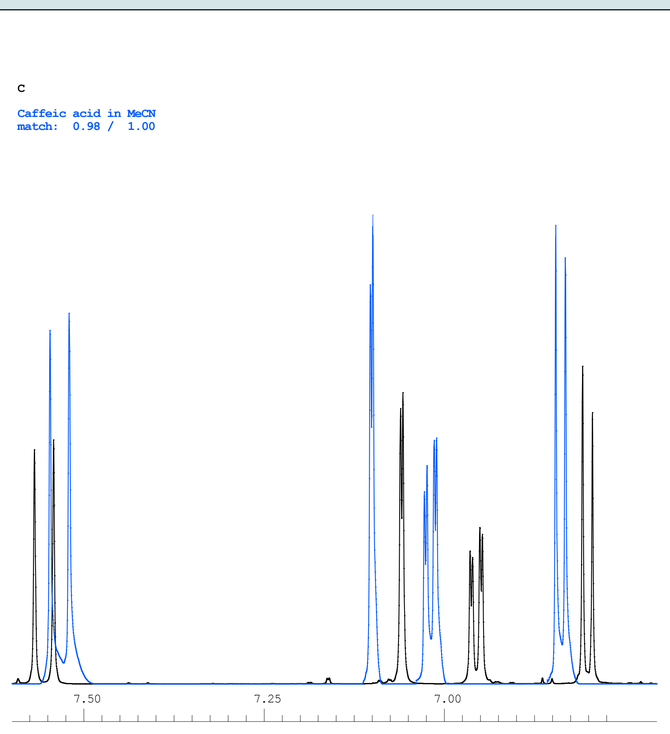
Database searches can either be conducted by superimposing the entire unknown spectrum on top of the database entries or through subspectra matches where individual signals are compared to the unknown, allowing small chemical shift changes.



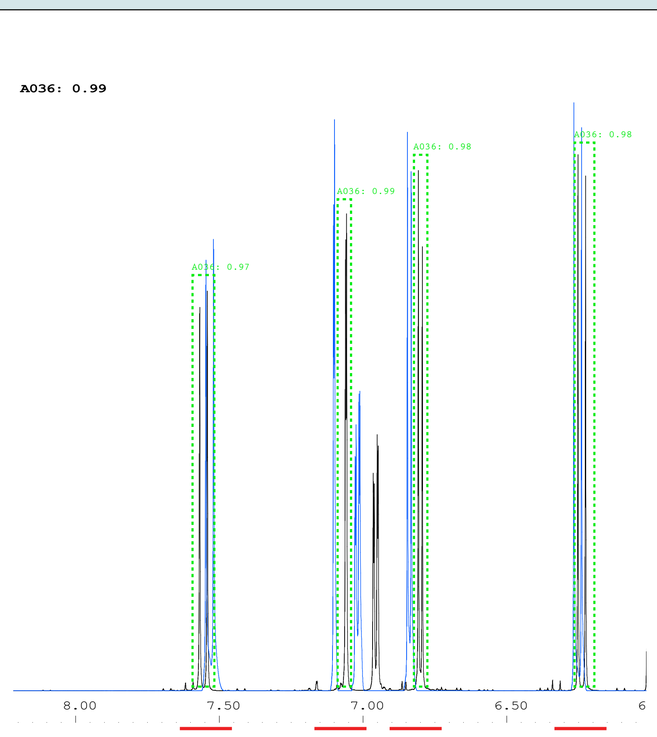
Through automation of the HPLC-SPE-NMR system it is possible to acquire high quality spectra of the majority of metabolites in an extract. By comparing the information-rich ^{1}H NMR data to a database of known compounds, the strength of the dereplication process is greatly enhanced as structure information is taken into account. This ensures that only unknown compounds of interest takes up valuable spectrometer time for time-consuming 2D experiments and subsequent structure elucidation.



When matching an unknown against a database a visual presentation of the match percentages appears as a table of colored dots - in this case two compounds are matched. Pressing the green dot superimpose the unknown and the hit for manual inspection



Since chemical shift values are dependent on factors such as solvent, concentration, and temperature, AMIX allows small chemical shift variations. Here, a 98% match is found even though each signal is shifted differently.



By expanding the database to contain information about each signals multiplicity, integral and chemical shift range it is possible to match against key signals. This opens up the possibility of seaching for substructures or compound classes.

Results and perspective

Through this proof-of-concept study we have shown that the AMIX platform is suitable as a dereplication tool utilizing ^{1}H NMR spectra. The matching ability positively identified both strychnine and caffeic acid amongst a database of 80 compounds. Through further development of the AMIX platform there is great potential for rapid and precise dereplication of complex mixtures analysed by LC-NMR techniques. Furthermore, signal-specific matching is believed to become a valuable tool for classification of compounds into compound classes or for targeted, class-specific analysis.

References

[1] Jaroszewski JW, *Planta* Med 2005; 71: 795-82
[2] Exarchou V *et al. Current Trends in Analysis and Characterisation* 2006: 143-155

Acknowledgements

The author would like to thank Bruker Biospin GmbH and the Drug Research Academy for co-financing the PhD fellowship. Both Dr. Christian Fischer, Dr. Klaus-Peter Neidig and Dr. Nils Nyberg are thanked for valuable help in building the database